

**SUMMARY** | Data scientist with 7 years of experience building complex, large-scale data pipelines and inferential and machine learning models. Experience running online A/B experiments, managing complex compute infrastructure and confidently presenting to VP-level audiences. Passionate about innovation, creativity, tech, data, engineering, leadership, mentoring and team building. Startup co-founder for 18 months.

## EXPERIENCE

### DATA SCIENTIST, ENGINEERING

#### GOOGLE, YOUTUBE MUSIC

APR 2021 - PRESENT

Built data-driven innovations that enhance music recommendations on YouTube:

- Designed a novel technique to **improve recommendations for music discovery for one billion users**. Used regression techniques and A/B experiments to estimate effects of improvements. Presented findings to VP-level audience. **Processed, combined, and intelligently sampled billions of rows** across 15 data sources of video metadata and event-level user logs. **R, SQL**
- Simulated alternative musician payment strategies using billions of user-song listening events and revenue logs. **R, SQL**
- Analyzed the impact of YouTube Shorts on music discovery and presented to Director-level audience. **R, SQL**
- Fast, independent learner. Became self-sufficient within three months of joining. Manager remarked, "I feel like you've been here for a year."
- Lead changes in YouTube Data Science tech infrastructure by developing R packages, templates, and documentation.

### DATA SCIENTIST, ENGINEERING

#### GOOGLE MAPS

MAR 2020 – APR 2021

Designed and implemented operational metrics to guide organization based on dashboards and insights:

- Systematically measured data quality** of billions of locations by producing a daily-updating dashboard used by 20+ engineers. **SQL**
- Helped iterate and experiment on new software features by reporting on **six A/B experiments** using large-scale online experimentation tools.
- Influenced product understanding with **ML-driven insights** across 15 presentations and reports. **Python, Regression, Clustering**

### MANAGER DATA SCIENTIST, VALUATIONS

#### CAPITAL ONE

JUN 2019 – MAR 2020

Rebuilt a five-year-old customer valuation framework from scratch in the cloud in six months using engineering best practices:

- Increased present value of application program by \$80M by training **12 machine learning models** to estimate the lifetime profitability of **millions of customers**. Implemented feature selection, hyperparameter tuning, and model validation techniques. Optimized across **decision trees and logistic regression** to select the best modeling methodology. **Python, XGBoost, H2O, Dask**
- Processed 20x more data (**1.6 billion rows, 1.7 terabytes**) by designing and developing a **data mining pipeline** that combines eight data sources using distributed computing techniques. **Python, pandas, Dask, SQL, Docker, Linux, EC2, Kubernetes**
- Coordinated across **20+ teams** to consolidate redundant technologies through rigorous prioritization and clear communication.
- Automated software packaging and delivery by implementing 15 code quality checks using **CI/CD**. **Python, Jenkins, Docker, GitHub**
- Supervised three data scientists** by providing daily feedback, mentorship, code reviews, and development opportunities.
- Evangelized learnings to diverse groups of stakeholders and community by sharing dozens of resources and providing multiple trainings.

### PRINCIPAL/MGR DATA SCIENTIST, RISK

#### CAPITAL ONE

JUL 2017 – JUN 2019

Developed an end-to-end data processing and machine learning classification pipeline to predict customer behavior in real-time:

- Designed and developed a **data mining pipeline** from scratch that queried, cleaned, and combined **2 billion rows from 17 tables** into a single view of customer behavior. Developed pipeline as a **Python** package with over **26,000 lines of code** complete with logging, configuration files, code quality (unit tests, code coverage, etc.), and command-line tools. **Python, PySpark, SQL, shell scripting**
- Trained **1000+ machine learning models** and identified top 65 predictors of customer behavior. Developed feature selection, model selection, and model validation methods. **Python, H2O, Docker, Databricks, decision trees, logistic regression**
- Simplified insight-generation process by building an interactive web application to visualize and compare model results. **R, R Shiny**
- Coordinated implementation of real-time model with key stakeholders, business analysts, and software engineers. Presented findings weekly.
- Promoted to Manager within one year of hire; management considered it a 'rare event', citing exceptional leadership and innovation. **JUL 2018**

### CONSULTANT II

#### BATES WHITE ECONOMIC CONSULTING

AUG 2015 – JUL 2017

- Saved colleagues **110+ hours per month** by developing 67 tools from scratch that were used over 20,000 times in six months. **Excel VBA**
- Forecasted **LIBOR** interest rates using 13,000 time series regression models and produced 30,000 plots to validate performance. **R, MATLAB**

### STARTUP CO-FOUNDER

#### NOVI SECURITY

JAN 2013 – AUG 2014

- Achieved **top 98.7% of all-time grossing** Kickstarter projects by selling \$175,681 of product to 848 customers in 30 days ([link](#)).
- Raised **\$560,000 of seed investment** by pitching business model to 15+ angel, venture capital, and private equity investors across four states.

# BRYAN WHITING

bryan.g.whiting@gmail.com | 203.822.4355

Python, R, SQL

[BryanWhiting.com](http://BryanWhiting.com) | [GitHub](#) | [LinkedIn](#)

## PASSION PROJECTS

## OPEN SOURCE CONTRIBUTIONS

NOV 2016 – PRESENT

- Scraped 3,500 articles, performed **text analytics**, and used **GitHub actions** to post insights to social media on a daily schedule. **R, Docker**
- Created five **open source data science blogs**, sharing demos in R and Python with 23+ posts and **20,000+ all-time page views**.
- Scraped 15,000 used car prices to research and identify undervalued cars and optimal buying prices (see [post](#)). **R, dplyr, ggplot2**
- Coded **Bayesian hierarchical model** (Gibbs sampler) to forecast **revenue** of [golf course tee times](#). **Python, R**

## LEADERSHIP, AWARDS, AND SERVICE

- **ACCELERATED TALENT MANAGEMENT, CAPITAL ONE** (2019): Selected as one of 30 high-achieving data scientists with leadership potential.
- **CONGREGATION LEADER** (FEB 2018-AUG 2019): Served 200+ members, conducted 15 personal interviews per quarter, and mentored youth.
- **CROCKER INNOVATION FELLOW, BYU** (JAN 2013): Selected as one of 20 university students to receive \$10,000 to pursue entrepreneurial ideas.
- Taught self-reliance principles to thousands across six cities as a volunteer, Spanish-speaking missionary in Honduras (AUG 2009-AUG 2011).

## EDUCATION AND SKILLS

EDUCATION	<b>Brigham Young University (Provo, UT):</b> B.S. in Statistics (APRIL 2015), <b>M.S. in Statistics</b> (APRIL 2015). MASTERS PROJECT: Built statistical framework to identify outliers in count data with class imbalance ( <a href="#">link</a> ).
PROGRAMMING	<b>Python:</b> <a href="#">H2O</a> , <a href="#">PySpark</a> , <a href="#">pandas</a> , <a href="#">Dask</a> , <a href="#">XGBoost</a> ; <b>R:</b> <a href="#">ggplot2</a> , <a href="#">Plotly</a> , <a href="#">Shiny</a> , <a href="#">dplyr</a> , <a href="#">tidyverse</a> ; <b>SQL;</b> <a href="#">git</a> ; <b>shell scripting;</b> with prior experience in <b>MATLAB; Stata; Excel VBA; C.</b>
CLOUD	<b>AWS:</b> <a href="#">EC2</a> , <a href="#">S3</a> , <a href="#">Redshift</a> , <a href="#">EMR</a> ; <b>Linux;</b> <a href="#">Databricks</a> ; <a href="#">GitHub</a> ; <a href="#">Jenkins</a> ; <a href="#">Docker</a> ; distributed computing
MACHINE LEARNING	Regression, classification, decision trees, statistical methods (frequentist, Monte Carlo, Bayesian).